Ethic Theory Moral Prac (2015) 18:223–235
DOI 10.1007/s10677-015-9574-8

Moral Hypocrisy and Acting for Reasons: How Moralizing Can Invite Self-Deception

Maureen Sie

Accepted: 24 February 2015 / Published online: 17 March 2015

Abstract According to some, contemporary social psychology is aptly described as a study in moral hypocrisy. In this paper we argue that this is unfortunate when understood as establishing that we only care about appearing to act morally, not about true moral action. A philosophically more interesting interpretation of the "moral hypocrisy"-findings understands it to establish that we care so much about morality that it might lead to (1) self-deception about the moral nature of our motives and/or (2) misperceptions regarding what we should or should not do in everyday or experimental situations. In this paper we argue for this claim by elaborating on a fascinating series of experiments by Daniel Batson and his colleagues who have consistently contributed to the moral hypocrisy findings since the late nineties, and showing in what way they contribute to a better understanding of moral agency, rather than undermine the idea that we are moral agents.

Keywords Moral hypocrisy · Self-deception · Acting for reasons · Moral agency · Daniel Batson

1 Introduction

Moral principles such as "you should alleviate suffering when possible" are regularly exchanged in explaining and justifying our actions and practices. Conversations about what to think of certain actions, measures, or policies often turn on finding principles people can agree on. Also a substantial part of moral philosophy is spent on arguing about which moral principles do regulate and should regulate our lives. Hence it should be no surprise that we expect people to conform to the moral principles they articulate. We positively dislike people

M. Sie (✉)
Department of Philosophy, Erasmus University, Rotterdam, The Netherlands
e-mail: sie@fwb.eur.nl

Maureen Sie
e-mail: m.m.s.k.sie@phil.leidenuniv.nl
URL: http://maureensie.wordpress.com/

M. Sie
Department of Humanities, Leiden University, Institute of Philosophy, Leiden, The Netherlands

who advocate moral principles but do not act in accordance with them and condemn those who act for so-called ulterior motives. "Moral hypocrisy" is the general label for this phenomenon.

According to some, contemporary social psychology could be described as a study of this phenomenon (Monin and Merritt 2012). Many experimental studies in social psychology have the familiar setup of showing that we cannot take proclaimed moral intentions at face value and that proclaimed moral principles only determine what we do and how we morally evaluate one another under very specific conditions. To give a few random examples: when the costs of acting in accordance with certain moral principles become too high, people readily let go of those principles (Batson and Thompson 2001). When faced with the opportunity to live up to intended moral actions, many people fail to live up to them (Epley and Dunning 2000), we evaluate ourselves and those who belong to our in-group more positively and leniently than we do others and those who do not belong to our in-group (Valdesolo and DeSteno 2007).

Many moral philosophers do not take note of, or are not impressed by, these findings. The reason for this might be that the existence of moral hypocrisy is not contentious. Anyone reading great literary novels or observant of human behavior already knows that we often do not practice what we preach and are biased in our judgments. This is perhaps the reason why "moral hypocrisy" researchers sometimes exaggerate their findings couching them in a vocabulary that suggest we are "nothing but hypocrites" and that we should worry even whether people are ever truly moral. These exaggerations notwithstanding, as scientists their aim must be to show something about morality in general. In this paper I argue that what they show about morality in general is: (1) how widespread and important the "desire to appear to act morally" is and, as a result of that, (2) that labeling things as "moral" or suggesting something is "the moral thing to do" has a huge impact on our self-understanding, self-reports and actions. That second insight should impact the research practices of those who study moral hypocrisy, but that is not what I will focus on. Rather the aim of this paper is to argue that contrary to how social psychologist tend to frame their findings, our desire to appear to act morally should not be perceived of as undermining or contrary to "true moral agency." That framing sticks to a portrayal of moral agency that fails to appreciate its complex and interesting nature so well captured by the "moral hypocrisy"-experiments themselves.

To get a better grip on the moral hypocrisy paradigm and be able to explain why it should interest moral philosophers, I start with a more detailed look at a series of fascinating experiments by one of the consistent contributors to the hypocrisy literature, Daniel Batson.

In the second section I argue that contrary to what Batson and his colleagues assume, it is not at all clear that people act morally wrong in the experiment. What is interesting, though, is that the participants to the experiment seem to think they are: they evaluate their own action negatively and they articulate a moral principle that they subsequently transgress. On the basis of that phenomenon, I argue that apparently we can be confused about what we should do in what circumstances and for what reasons. The Batson findings suggest that we sometimes even *deceive* ourselves about the nature of our motivation as a result.

In the third section, I articulate a view on moral agency that allows us to understand how we can be confused about what we should do in what circumstances and for what reasons. I start by defining the concept of "reasons" in such a manner that (1) it does not decide on any of the many controversial issues in philosophy that surround this concept and (2) enables us to make sense of the distinction between explicit/articulated reasons for our actions and non-explicit/unarticulated ones. Many of the things we learn we learn by participating in moral practices, not by explicit articulate instruction. As a result, I argue, identifying and/or articulating the principles that regulate or should regulate our everyday dealings is not simple and straightforward as the Batson (and other) experiments presuppose.

In the fourth section, I return to the Batson experiment with the distinction between articulated and non-articulated reasons in mind. I argue that a much more interesting interpretation of the findings is available once we abandon the assumptions that it is an easy matter to identify and/or articulate the relevant moral principles in every situation and that it is clear what it means to identify or affirm such a moral principle. I conclude that what the experiments show is that articulating a principle as "moral" suffices to influence people substantially even if we do not and should not accept that principle in our everyday practice. This is in line with the observation of Batson and his colleagues that we care enormously about appearing to act morally, but abandons, as I will explain, their narrow interpretation of what it means to truly act morally.

2 The Moral Hypocrisy Experiment of Batson et al.

The participants of the Batson experiment are told they are taking part in an experiment on the relation between task performance and reward. The experimenters tell them there are two tasks: one boring and tedious task with no reward attached (NEG) and one easy task that earns you a raffle ticket when performed adequately (POS). The raffle ticket gives participants a chance of earning 30 dollars at the end of the experiment on top of their payment for participation. Beside the main aim of the experiment (investigating the relation between task performance and reward), the experiment also tests an additional hypothesis. Because of that additional test, so the participants are told, they are also asked to divide the two tasks (POS and NEG) between themselves and another participant. They are told that the other participant will not know that another person decided on the division of the tasks, they will not meet this participant, and they are allowed to make their choice alone and behind closed doors. What do people do and how do they evaluate their choice with hindsight?

Daniel Batson and his colleagues show that 16 out of 20 persons "will take the POS task" (80 %), but rate their choice as "not very moral" (Batson et al. 1999, 2002; Batson 2008).[1] That is, on a scale from 1 to 9, they rate themselves with a 4.4.[2] The participants who "took the NEG task" rate themselves with an 8.4. Most of us apparently do not act as we judge morally best. What interests the researchers is not the fact that we do not act in a morally exemplary way by our own standards; instead, the researchers are interested in the question of whether moral hypocrisy—wanting or desiring to appear moral—is a better explanation of our choices in the experimental setting than wanting to be truly moral. According to their view to act truly morally would be to act in accordance with a moral principle one accepts.

In order to make an experimental investigation of their question possible, they (1) make explicit a principle they found to be effective in prior runs of the experiment. They tell the participants that in a former run of this experiment, people indicated that they thought "it most fair to give each an equal chance on the positive task" (hereafter, we refer to this as "the equal chance" principle, or ECP for short), and that therefore they are given a coin packed in a sealed plastic wrapper.[3] If they wish they can use this coin, so the participants are told, to divide the tasks in a fair way. Again, the participants are left alone behind closed doors to make their choice, either with or without using the coin. As will become clear, the coin adds (2) elbowroom to cheat.

[1] Note that the size of the groups participating in these experiments is often small and also that the results listed concern group effects and mean responses.

[2] One and nine represent, respectively, "my choice was 'not morally right' and 'morally right.'"

[3] Unless otherwise stated, we will understand "moral principles" broadly, that is to include what moral philosophers might be more inclined to refer to as "behavioral principles." I will have something to say about this loose use of the concept of moral principles in §2 and §4 of this paper.

So what did the participants do? Fifty percent use the coin and the other 50 % do not. Interestingly, in *both* cases, 90 % of the people end up assigning POS to themselves (hereafter: "take POS"). Moreover, those who end up with the POS task after having used the coin rate the moral nature of their choice with a 7.11; those who did not use the coin rate the moral nature of their choice with a 3.56 (more about this below). Since we know coins do not cheat, we can infer that part of the participants made use of the available elbowroom, that is, they take POS regardless of the outcome of the coin. This was affirmed in a follow-up of the experiment in which Batson and his colleagues found that those cheating—those who did not observe the outcome of the toss—rated their action as more moral than those who assigned themselves the positive task without cheating.

The results of this experiment remain more or less the same in slightly different setups. When one adds serious negative consequences to inadequate task performance, e.g., administering electric shocks, people will feel less inclined to use the coin (30 %) and will take POS even more often (100 %). When it is communicated that the other participant will be informed about how the task assignment was determined, more people will use the coin (80 %), but the division of tasks remains the same. Hence, as the researchers conclude, if we can appear to be moral without paying the actual price of acting in accordance with the moral principle made salient to us, many of us will do so. That is, if it is communicated that the other will remain in the dark about how the tasks were divided, around four out of every five participants will take POS even though they acknowledge that the most moral choice would be to give each an equal chance to get the positive task. When we do decide to use the coin, a substantial part of us will take POS regardless of the outcome of the coin. Moreover, as previously mentioned, the collective of those who toss the coin perceive their action (cheating or not) as "more moral" (7.11) than the collective of those who do not (3.56).

Let me make explicit what I take the experiment to establish thus far: around one out of every four participants decides to toss the coin and takes POS, among those several do so regardless of the outcome of the toss.[4] Regardless of the fact that some of the participants cheat (take POS regardless of the outcome of the toss), the collective of those who flip the coin feel morally superior to the collective of those who do not. Let me call the fact that the ones who toss the coin end up with POS in much more than 50 % of the cases the "coin discrepancy effect" (CD effect) and the fact that the collective of participants who toss the coin understand their action as "more moral" than the collective that did not toss the coin the "moral superiority effect" (MS effect).

Batson and his colleagues' research question is how exactly these participants are deceiving themselves. Do they fool themselves (1) into believing their choice is the moral one by fiddling with the coin and not paying attention too closely to which sides comes up? Or (2), does the deceit involve a failure to compare their own choices with the behavioral ECP they explicitly adopt? To check on the first possibility, Batson et al. repeated the experiment with clearly labeled coins, "POS to Other" on one side and "POS to Self" on the other. This took away the ability to fool oneself by not paying attention too closely, e.g., by fooling oneself about which side of the coin came up or by fooling oneself about which side was attached to which task. The labeling did not change the results. It did enable them to establish the aforementioned MS effect. By secretly observing the participants[5] they could divide the participants into several groups, one group of which flipped the coin and cheated with the result. This latter group rated

[4] We assume that the coin is fair and assigns "POS to Self" one out of every two tosses. When half of the participants toss the coin, then one quarter of all participants actually get the "POS to Self" assigned in a fair manner. The other quarter, or slightly less, cheats.

[5] This was made possible by using colored labels distinguishable through a tiny opening in the paper covering a small window to the room in which the participants assigned the tasks.

themselves with a 5.56, which is higher than the group who assigned POS to themselves without flipping the coin (3.89) (Batson 2008, 59–60; Batson et al. 2002, study 2).

To check whether the second kind of self-deception is involved, Batson (2008, 61) repeated the experiment under conditions of heightened self-awareness. In the seventies, experiments on cheating that disclosed heightened self-awareness reduced cheating behavior and could easily be manipulated by the presence of a mirror (Diener and Wallbom 1976; Wicklund 1975).[6] When the participants made their choice in the Batson experiment with a mirror hung on one of the walls, the CD effect disappeared completely. And of those who did not use the coin, the amount that took POS themselves dropped significantly (to 62 %).

And it gets more interesting. When ECP was not made salient prior to the choice and self-awareness heightened, this partly re-installed the CD effect. In addition to this, it lead the participants to adopt a principle different to anything found in prior trials, i.e., a principle that to take the POS yourself is actually the morally correct choice (Batson et al. 1999).[7] Batson and his colleagues (1999, 533) take this as further proof of moral hypocrisy. When we cannot get away with cheating due to heightened self-awareness we will pick a standard to suit our actual choices. Without the heightened self-awareness we cheat by not comparing our own actions with the standards we set ourselves, we change strategy and articulate a standard that makes our choice come out as the moral one, after all. I will come back to this interpretation in the last section. First, let me go into the general picture of moral agency that the researchers take for granted and that is disclosed by two contentious assumptions of the experiment.

The first assumption is that we only act "truly moral" when we act in accordance with moral principles made explicit or affirmed beforehand. This, for a variety of reasons, seems a very narrow interpretation of what it means to act morally. Strongly related to that is the second assumption, i.e., that it is a trivial and unchallenging matter to identify and/or articulate which moral principles should be applied in a certain situation. The experimenters assume that all we have to do is ask people what we believe would be the most moral thing to do, and we get as answer the moral principle they accept. However, as I will argue, their findings—when properly understood—strongly suggest that this is not how it works at all. I criticize the two assumptions in Section 3, in the next section let first me discuss ECP.

3 The Equal Chance Principle

In our daily lives we constantly have to balance pursuing our own interests with doing the morally exemplary thing: buy a silk blouse we do not really need or donate to a charitable organization, ask the neglected but very difficult child we feel sorry for over to play with our daughter or not, be patient and kind with those around us or do what is demanded of us in pursuit of our interesting jobs, and so on. When asked whether we could have performed better "morally speaking," many people will say that they "could have" when they did something that benefitted themselves at the expense of others. This, however, does not mean that we believe we transgressed a moral principle "in the strict sense" in those instances, just that we think we could have performed better morally speaking. Something similar seems to be the case in the Batson experiments. Intuitively there is something that "speaks in favor of giving

[6] Contemporary findings in very diverse fields have found, much in line with this, that heightening self-awareness by the use of 'eyes' influences moral behavior as well.

[7] Of the 14 participants in the low standard salience condition and high self-awareness, four persons stated that the most moral thing to do was to assign the positive task to oneself, five that there was no morally right way to divide the tasks, three that one should use a random method, and three that one should assign the positive task to the other participant. Batson et al. (1999).

each an equal chance at POS" but when we do not it does not seem to be the case that we act wrongly, transgressed a moral principle or failed to live up to an obligation. To see what exactly it is that speaks in favor of ECP, why people articulate ECP in the experimental situation, and whether we do or do not accept it as a principle in the strict sense in our everyday practices, let us consider some situations resembling the experimental one.

Compare two scenarios. Scenario (1): two people simultaneously kneel down to help someone pick up the groceries that fell on the boardwalk. At that instant, both see a very nice and expensive-looking bracelet within arm's reach (hereafter, S1, or Similar Position). Scenario (2): one person kneels down to help someone pick up the groceries, another one who was standing a bit further away hurries to the scene to do the same. Before the other person arrives at the scene, the person kneeling down sees a very nice and expensive-looking bracelet within arm's reach (hereafter, S2, or Dissimilar Position).

It seems that ECP suggests itself naturally, but only in S1. When I kneel down to help someone and see a nice bracelet, I am under no obligation to share my good luck with people arriving at the scene a moment later. Moreover, even in S1, ECP suggests itself only because both people see the expensive-looking bracelet *simultaneously* and are in no position to deny it. When, for example, a large carton of cereal obstructs the view of the bracelet for the other person, ECP will not suggest itself. "Now look what *I* found here," is an obvious thing to say in that scenario (S1a, Obstructed View). In this situation, it is also clear that you are the lucky one to have kneeled down with a clear view on the bracelet. Perhaps, the situation changes when the person kneeling down simultaneously is your best friend (S1b, Good Friend).

So what distinguishes these scenarios from one another? The most salient difference between Similar Position (S1) and Dissimilar Position (S2) seems to be that in S1, grabbing the bracelet would be awkward because it fails to acknowledge the other person, i.e., to acknowledge the other person's "equal claim to the object." The equal claim to the object derives from the fact that both of you are in *exactly* the same position and cannot plausibly deny that this is the case. It seems that capturing another person's gaze plays an important role in our everyday competition for, for example, empty tables in a restaurant or parking spots. Looking straight into one another's eyes makes it difficult for us to deny being in exactly the same position with respect to the desired object (table, parking lot). This seems to be why Obstructed View (S1a) provides us some leeway to bluntly claim the object for ourselves. In this case, we can plausibly deny that the other is in exactly the same position as we are. It is also why it would be awkward to do so when the other person is your friend, as in Good Friend (S1b); after all, friends are supposed to share their good fortune regardless of a small obstacle like a box of cereal.

Of course it would be kind and praiseworthy in all the above scenarios when the person kneeling down first offers to give the other person an equal chance on the bracelet (ECP)—e.g., proposes to throw a coin to decide who gets it. However, that does not mean that we are morally required or believe to be required to do so. When we take a close look at our everyday practices we do not seem to feel a need to distribute our luck among those close by or to decide on a procedure for an equal distribution unless the other person is indisputably in an equal position. Let us return to the Batson experiment.

In the Batson experiment, the participant who is so lucky to arrive "first at the scene" is allowed to decide on the division of tasks and has no contact with the other participant. So is the "dividing participant" in an equal position as the other participant? Bearing in mind the distinctions between the slightly different scenarios, the experimental situation is clearly ambiguous. The fact that the experimenters raise a question about "the most fair thing to do" or in subsequent experiments communicate ECP suggests to the participants that the other participant is in an equal position. After all, why would fairness or ECP matter when the other

is not in an equal position? On the other hand, the other participant is nowhere near in sight and in most setups of the experiment, will not be met afterward either. Moreover, it is communicated to the dividing participant that the other will not know she/he divided the tasks. Hence, in this respect, the dividing participant is clearly the lucky one with ample leeway to decide whether or not to share her/his good fortune with the other participant. As a result of that interpretation of the situation, we cannot infer that the participants who took POS did anything morally wrong, even according to their own standards.

The observation that ECP is not unconditionally applicable in the experimental situation changes the interpretation of the experiment drastically. Rather than showing that we only act morally when it is absolutely required to appear to act morally, the experiment rather seems to show that articulating a principle "as moral" suffices to influence us substantially even when we do not accept that principle in our everyday practices. It influences how people morally evaluate themselves, what they will do under conditions of heightened self-awareness, or when altruism is triggered; for some, it even leads to self-deception (CD effect and MS effect). Before we discuss this alternative interpretation more elaborately, let us explain why it makes sense to think that it is not a trivial and unchallenging matter to identify and/or articulate which moral principles are and should be applied in situations such as the experimental one.

4 Reasons: Efficaciousness and Explication

When we are asked why we did something in an everyday moral setting we will regularly allude to one or another moral principle or consideration. Why do you never ask over that difficult and neglected child? Because I am primarily responsible for the wellbeing of my own children! Many of these answers satisfy us. That is, we think they justify the action and assume the articulated considerations (moral principle, reason), in one way or another, figure in the explanation of the action. *Prima facie* this might give rise to a picture of moral agency as is presupposed in the Batson experiments: a picture in which agents act "truly moral" when the principles and reasons they articulate are those they actually act in accordance with. However, that picture is overly simplistic. Before I explain why, let me define reasons in a way that does not decide on the many discussions in philosophy that revolve around that notion and allows us to make sense of the distinction between "articulated and explicated reasons" and reasons we are not aware of. As I explain later on, it is this distinction we need in order to make sense of the Batson experiment and experiments like it.

As a definition that suits all in the sense that it does not decide on the many controversial issues that revolve around it, we could say that reasons are "*justifying* explanatory states." That is, states that are part of a larger network of states; a network with a certain aim. In this very broad definition, almost everything can function as a reason as long as it is taken as part of a larger network of states. According to this definition, it makes no literal sense to say of someone tripping on a carpet, let us call her "June," that the reason she hurts herself is the curling carpet. A more apt description is that the curling carpet *causes* her to trip. It does make sense to say that a cat chases a mouse because of her hunting instinct for food, for this hunting instinct is an explanatory state that is part of a larger network of "survival" states that justify hunting a mouse when, e.g., a cat is not fed by its human caretakers. However, it makes only metaphoric sense to say of a cat that she spoils the carpet because she wants to catch the sunlight (damaging it with her long nails chasing reflected rays of sunlight). In this case, the cat's attack of the carpet is not a possible justifying explanatory state; it is not part of a larger network of states with a certain aim. The cat is not spoiling the carpets for a reason; her playfulness and her long nails cause the damage to the carpet. Then again if the cat is throwing

up because she ate something toxic, it does make some sense to say that she is throwing up for a reason, because in this case, the throwing up is part of a network of "survival" states. It seems to me that this basic, trivializing distinction between reasons and other explanatory states[8] identifies an important distinction in our everyday discourse, i.e., "events that must be made sense of" and "events that lack such overarching sense."

Within that very broad definition of reasons, an entity acts for a good or bad reason if the state that explains her or his action is part of a larger network of states with an aim that *does* or *does not* justify the action. A good reason is a state that justifies the action in view of the overarching aim and a bad reason is one that does not. For example, the cat acts for a good reason if the mouse is a real one, but not when it is a mechanical one or when she gets enough food from her caregiver. The kind of mistake the cat makes is different in each case, and we can probably make up many other ones.

As a consequence of our definition of reasons, we can understand "our ability to act for reasons" as our ability to act or refrain from acting on explanatory states that might justify it. Hence, in so far as we believe that a cat could have refrained from acting on her hunting instinct, for example, because she fears the anger of her human caretaker, we believe the cat has the ability to *act for reasons*.[9] Note that it is a characteristic feature of this very broad and basic understanding of reasons that we can act for reasons without being aware of it in the sense that we recognize that we acted for them and are subsequently able to report on them when asked.[10] There are many reasons that explain our actions—good or bad—without us acknowledging them as such, or us even being aware of them. Nevertheless, these reasons do make sense of what we do or do not do. There is a whole network of considerations that regulates and explains our behavior and actions, even though we are not aware that it is doing so and we might never have been.

In line with that broad definition "moral reasons" can be defined as morally justifying considerations, i.e., considerations that are part of a larger network of moral considerations—a network with a certain overarching aim that concerns our relation with a group of other beings or the world. This broad definition enables us to see what is wrong with the first of the assumptions we observed in Section 1, the idea that we only act "truly moral" when we act in accordance with moral principles made explicit or affirmed beforehand. Morality plays an enormous role in our lives. From an early age onward, we are constantly morally evaluated and responded to with moral sentiments such as blame, resentment, moral indignation, praise, and gratitude. We are raised to do and not do certain things, made to explain and justify ourselves whenever we transgress certain normative expectations, disliked or liked for how we behave, and punished or lectured for harming others or inconsiderate behavior. As a consequence, moral agency—broadly understood—should be pictured along the lines of our participation in traffic (cf. Sie 2014). Much of what we learn and pick up on in the moral domain escapes and/or precedes "deliberative awareness." We are not instructed or explained how to do it, in what way, and why: we learn it in practice. Of course we also learn a lot of rules of thumb, learn

[8] An explanatory state is a state of an entity that explains her/his particular actions or movements, i.e., in the absence of these states, that movement or action would not have occurred. The cat would not have spoiled the carpet if her nails where clipped, June would not have tripped if the carpet had not made her lose her balance. Playing while having long nails and loosing one's balance are explanatory states, but not possible justifying one's in so far as they are not part of a larger network of states with a certain aim.

[9] Cf. MacIntyre (1999). To claim that cats and other animals are able to act for reasons is not meant to decide on the issue of whether human ways to act for reasons are fundamentally distinct from animal ways to act for reasons.

[10] This is the reason why I am deliberately not differentiating between, for example, acting on a reason and acting *in accordance* with reasons, where the first but not the latter is easily understood as implying an awareness of the reasons.

what is morally right and wrong by being told so, and by being explained why certain things are considered right or wrong. However, we also learn what is morally wrong and right by participating in social practices and institutions. By such participation we also learn how to apply our "moral knowledge," as we do in our interactions with significant others such as our family, friends and role models. Hence, our ability to function adequately in the moral domain is partly constituted by learning how to respond to what and whom, at which junctures and signals, and to do so automatically without prior reflection.[11] Hence, we may act in a morally adequate manner without being able to identify and/or articulate the moral principle that regulates our actions. As a consequence, we cannot equate "true moral action" with "action in accordance with moral principles articulated prior to our action," which is what Batson and his colleagues do. Nor can we simply assume that people know and/or are able to articulate for what moral reasons they act.[12] Let me explain these last claims.

Humans are emotionally and motivationally complex beings who tend to act and evaluate on a plethora of motives, beliefs, and circumstances. On many occasions a whole set of diverse considerations and overarching aims comes into play, the complex combination of which make us respond in the way we do. Annoying noises wake us up (the alarm clock goes off), our children need to go to school, we have a meeting, we feel energetic and want to get active, we promised a colleague we would see her at work, and so on. Why did we get up? Because the alarm clock woke us up? Because we love our job, are responsible parents, keep our promises? Most of the time the adequate answer will consist of a combination of considerations only few of which we are aware of and/or care to mention when asked. We are not fully aware of, do not think and reflect on, all the overarching frameworks that make sense of the myriad of triggers we respond to (alarm-clocks, promises), things we do (get up everyday, take the car to bring the children school) and projects we pursue (become a philosopher). Many of them we take as given and legitimate. It is when conflicts or tensions arise or when we are confronted with moral demands or evaluations, that we articulate our moral reasons or principles in response (cf. Sie 2014). And in many cases the reasons we articulate, we articulate without much prior thought too, focusing on those we have learned to be or think appropriate for the occasion.

We are thoroughly embedded beings from the moment we are born, surrounded by devices and artifacts, involved in all kinds of institutions, personal relationships, and activities, many of which we did not choose or decide upon but that constitute our lives nevertheless. Unless we are on vacation or enjoy a free weekend, we do not typically get up in the morning, make up our minds about what to do, why and how to do it, and control our actions in light of what we decided or intended to do. Let alone that there is a moment in time at which we do this for our future as such. In a very similar manner our evaluations, values, and normative expectations are not neatly organized, thought through, or clearly articulated at one point in time. That does not mean that we never act morally or that moral considerations do not matter for us. It does mean that we do not always know what moral considerations matter to us and make us act as we do, and might even be mistaken with regard to them. With this in mind, let us return to the Batson experiments and the two assumptions articulated at the end of Section 1.

[11] Hence I am in full agreement with Nomy Arpaly (2003) who has argued, that we sometimes act for moral reasons without being fully aware of it or even without knowing it.

[12] By now several philosophers have developed accounts that explicitly accommodate the insight that we do not always know for which reasons we act, partly in response to the developments in the behavioral, cognitive, neurosciences that have brought the automaticity and the impact of influences that escape deliberative awareness to the fore. See for example Pettit (2007), Railton (2014). Although mainstream philosophy rarely discusses non-deliberative moral action, not all more traditional views are incompatible with it. See for a defense of that claim, Sie (2009)

5 Moral Hypocrisy or Overpowered Integrity

In my interpretation of the experiments, Batson and his colleagues show us that articulating a principle "as moral" suffices to influence people substantially even if they do not and should not accept that principle in their everyday practices. For that interpretation to make sense, one of two things must be the case: (1) The people in the experiment are mistaken about ECP (to give each an equal chance at POS) as the relevant moral principle or (2) they are mistaken about what it means to accept a moral principle like ECP. Given the nature of our everyday practices and the way we respond to reasons as set out in the previous section, both might very well be the case. What are the implications of either possibility?

When the first is the case, this would also allow for a re-interpretation of the follow-up experiment in conditions of high self-awareness without prior salience of ECP. In this experiment, some participants came up with a principle to fit their choice rather than ECP (Batson et al. 1999). Batson and his colleagues interpreted this as further evidence of our hypocritical nature. However, if ECP is not a principle that actually regulates our practices, rather than showing that we confabulate principles to suit our actual choice this follow-up experiment might show that conditions of heightened self-awareness accommodate a more honest and/or accurate perception of the relevant moral principles. There is some research to back up this interpretation.

There are some findings in social psychology that suggest that asking people their *intentions* in moral choice situations triggers "idealized" answers, answers about what they would do in an ideal world (bring home the difficult but neglected child everyday, not buy expensive luxury items but give more to charity, take NEG). When you pose two questions to people, i.e., one about what they would do in an ideal world and one about what they think they will actually do, they are much better, i.e., less hypocritical, in predicting their actual behavior (Tanner and Carlston 2009; Monin and Merritt 2012, 176). In one reading of these results, the reason for this is that people answer the question about what they would do in moral choice situations in terms of what they care about, underestimating the amount of other interests, available time and resources and so on, when push comes to shove. Something similar might happen when you ask people what "the moral thing to do" would be without raising their self-awareness, i.e., without making them pay attention to what they will do. If this is the case, adding the question "what do you think *you* would do in *this* situation regardless of what you think should be done in an ideal world" might actually make people adequately predict that they will take POS.

When the second possibility is the case, the participants of the Batson experiments understand the question of what they believe "they should do" as one about what is the most exemplary or excellent thing to do, not about what everyone should and would do under regular circumstances and/or should and would disapprove of when they do not. It is understandable that ECP suggests itself when asked what seems to be the right thing to do in the situation in this loose sense. First of all, as observed in Section 2, posing the question like that suggests that a *moral* standard is applicable in the situation. Secondly, the fact that another participant is involved who is affected by the choice you make contributes to the suggestion that the applicable moral principle must have something to do with fairness in division. Hence, it is no surprise that people articulate ECP. When this is the case, if people understand ECP as a moral principle in the loose sense of the word, it is interesting that it still brings some people to deceive themselves (CD effect and MS effect) and to act in accordance with it under conditions of heightened self-awareness or altruism. It shows that under the right conditions we are able to rise above ourselves, so to speak. That is we are made to feel sympathetic with other people or become very

aware of ourselves we might even go the extra mile when moral considerations are in play.

In any case, both of these interpretations do not give us cause to worry about morality in the sense that Batson and his colleagues suggest we should worry, i.e., that we do not really care about morality itself, but only about appearing to act morally. Rather, what their findings suggest is that we care so much about appearing to act morally that principles influence us even when, on second thought, we do not seem to accept them as moral ones in the strict sense, or might even reject them altogether. They influence us so much that it leads some of us to deceive ourselves.

Interestingly, as Batson and his colleagues discovered in another unpublished follow-up experiment, the kind of self-deception involved, rather than being hypocritical, is more adequately described as a case of what they call "overpowered integrity" (Batson et al. 2000). It looks as if people set out to act in accordance with ECP, but when push comes to shove, they fail. In this follow-up experiment, the participants were offered an additional decision option: they could allow the task assignment to be determined by the experimenter's flip rather than their own (Batson and Thompson 2001). Of those who decided to use the coin, 80 % chose to have the experimenters flip the coin, which indicates, as the researchers pointed out, that their initial intention is to observe ECP. This dropped to 25 % when the costs of observing ECP increased by telling the participants that ill performance on the negative task would be punished with mild but uncomfortable electric shocks. In this last version of the experiment, 50 % just assigned the POS to themselves, giving up, as Batson puts it, "any pretense of morality" (Batson and Thompson 2001, p. 56).

In the article in which they describe the follow-up experiment by Batson et al., Batson and Thompson (2001) conclude with a section entitled "Cost-Based Justification For Setting Morality Aside." They point out that the fact that we set aside morality whenever there is a personal cost is "tantamount to having no real principles at all." (Batson and Thompson 2001, p. 56). This shows that they remain true to the "moral hypocrisy" tune that is so popular in social psychology. This is a pity, since there is little reason to doubt that we often act hypocritically; therefore, this is not the most challenging outcome of their findings. Secondly, the idea that morality has to be pitched over and against our self-interest is superficial and confirms a rather simplistic picture of morality. The strength of their findings is that they show us to be (1) sensitive—over-sensitive we should say—to principles labeled as "moral," which leaves some of us (2) susceptible to self-deception of a very peculiar and sophisticated kind, i.e., to make us feel that we acted morally without acting in accordance with the principle made salient as the moral one (the CD effect). Besides inviting self-deception in certain conditions, articulated salient moral principles also make a difference on how:

(a) some of us act all the time (i.e., those who act corresponding to ECP);
(b) all of us act under conditions of heightened self-awareness (or when asked to imagine how the other would feel when they are assigned the NEG task); and
(c) we evaluate our actions (i.e., we evaluate them as less moral if we fail to act in accordance with ECP than if we succeed, more moral if we make ourselves believe we acted in accordance with the EC).

Therefore, we can conclude that presenting principles as "moral" seems to play an enormous role in our everyday lives regardless of whether (1) on reflection we would accept them, or (2) the principles do in fact regulate our everyday dealings with one another.

What does not follow is the conclusion that we only act morally when it is required to appear to act morally, unless we narrow down what it means to act morally to acting in accordance with moral principles we affirm *prior* to so acting.

Hence, as Batson and his colleagues rightly phrase it, their findings establish that we care a lot about appearing to act morally. This is probably why articulating and exchanging moral principles is so important and worth our while. When people usually act on a plethora of motives many of which escape their attention the fact that they care about appearing to act morally enables salient moral principles to trump other considerations even when we lack the time to thoroughly reflect on what we should do. Unfortunately that might sometimes backfire in the sense that it can lead to people deceiving themselves about the motivational origin of their actions. Therefore we should heed moral hypocrisy not only by being suspicious of people's actual motivations, but also by examining whether the principles we articulate are actually ones that regulate our daily practices or that we should want to regulate our daily practices.

6 Conclusion

Batson and his colleagues are right: we care a lot about appearing to act morally. We care so much, we have argued, that articulating moral principles will influence our actions even when on second thought we might not accept them. Given the kind of beings that we are and the important role that morality plays in our lives, this should perhaps not surprise us. We constantly have to act, decide, judge, evaluate and to balance ourselves in a network of normative expectations—what is expected of us as a student, parent, neighbor, colleague, and as a human being. Meanwhile, we have to respond to our immediate environment, which is partly constituted by our prior commitments and special relations, but is sometimes in tension with these. We constantly have to balance our own interests over and against the interests of those we live with and/or the things that matter to us and the things we desire. This, and the way in which we do it, is what moral agency consists of. We discuss and articulate clear moral principles to make sure this everyday trafficking will not cause major accidents and all the damage and misery that goes with them. But the way in which we apply these principles might be much more complicated than we realize or are able to articulate. When navigating our way through the complexities of everyday life, our eye is on our destinations (in the trivial everyday sense of that word, i.e., getting to work, making the deadline, and so on) and on how to get there. The travelling is largely automatic even though it is also regulated by complex social and moral reasons and principles. Social and moral reasons and principles that we are taught from an early age onwards and not exclusively by explicit instruction and articulation of those reasons and principles. As a result of the automaticity and the fact that we might not be aware of all moral reasons and principles that we act in accordance with, it can happen that we misidentify moral reasons or principles, as I have argued is the case in the Batson experiments. We might misidentify the principle that is applicable in a situation—as is the case for ECP—or the conditions in which the principle is applied, or we might misunderstand what it means to accept such a principle. To decide which of these is the case with respect to ECP, more empirical research is needed. What the Batson experiments do establish is that our caring to appear to be moral might lead to self-deception. Self-deception, so much we might agree on, is something that disables us to efficiently redirect our lives on the basis of our evaluation. Hence, one of the take-home messages of the moral hypocrisy literature from social psychology is that we should take care with moralizing.

Acknowledgments I thank the anonymous referees of the journal, Bryce Huebner, Philip Robichaud, Filippo Santoni De Sio, Nicole van Voorst Vader Bours and Arno Wouters for very valuable comments and discussion of earlier versions of this paper that enabled me to shorten and strengthen it a lot. I thank the editor of this volume, Katrien Schaubroeck, for her close reading of the paper and her helpful suggestions for improvements and Huub Brouwer for his editorial assistance. I thank the Dutch Organization of Scientific Research (NWO) that financed the research-project from which this paper is the result.

References

Arpaly N (2003) Unprincipled virtue: an inquiry into moral agency. Oxford University Press, Oxford

Batson CD (2008) Moral masquerades: experimental exploration of the nature of moral motivation. Phenomenol Cogn Sci 7(1):51–66

Batson CD, Thompson ER (2001) Why don't moral people act morally? Motivational considerations. Curr Dir Psychol Sci 10(2):54–57

Batson CD, Thompson ER, Seuferling G, Whitney H, Strongman JA (1999) Moral hypocrisy: appearing moral to oneself without being so. J Pers Soc Psychol 77(3):525–537

Batson CD, Tsang J, Thompson ER (2000) Weakness of will: counting the cost of being moral. Unpublished manuscript, University of Kansas, Lawrence

Batson CD, Thompson ER, Chen HJ (2002) Moral hypocrisy: addressing some alternatives. J Pers Soc Psychol 83(2):330–339

Diener E, Wallbom M (1976) Effects of self-awareness on antinormative behavior. J Res Pers 10(1):107–111

Epley N, Dunning D (2000) Feeling "holier than thou": are self-serving assessments produced by errors in self or social prediction? J Pers Soc Psychol 79:861–875

MacIntyre A (1999) Dependent rational animals. Carus Publishing Company, Chicago and La Salle

Monin B, Merritt A (2012) Moral hypocrisy, moral inconsistency, and the struggle for moral integrity. In: Mikulincer M, Shaver P (eds) The social psychology of morality: exploring the causes of good and evil. American Psychological Association, Washington

Pettit P (2007) Neuroscience and agent-control. In: Ross D, Spurrett D, Kincaid H, Stephens GL (eds) Distributed cognition and the will: individual volition and social context. MIT Press, Cambridge, pp 77–91

Railton P (2014) The affective dog and its rational tale. Ethics 124(4):813–859

Sie M (2009) Moral agency, conscious control, and deliberative awareness. Inquiry 52(5):516–531

Sie M (2014) Self-knowledge and the minimal conditions of responsibility: a traffic-participation view on human (moral) agency. J Val Inq 48(2):271–291

Tanner RJ, Carlston KA (2009) Unrealistically optimistic consumers: a selective hypothesis testing account for optimism in predictions of future behavior. J Consum Res 35(5):810–822

Valdesolo P, DeSteno D (2007) Moral hypocrisy: social groups and the flexibility of virtue. Psychol Sci 18:689–690

Wicklund RA (1975) Objective self-awareness. Adv Exp Soc Psychol 8:233–275